Mechanism Design

CMPUT 654: Modelling Human Strategic Behaviour

S&LB §10.1-10.4.2

Lecture Outline

1. Recap & Logistics
2. Mechanism Design
3. Quasilinear Mechanism Design

Logistics

- **Midterm** is this **Thursday, Feb 14**, in the usual classroom at the usual time
- **Assignment #2** will be released on **Friday**
- You should have received **solutions** for assignment #1 by email
- You should receive an email about **Mid-term Course and Instruction Feedback** either **today or tomorrow**

Recap: Social Choice

Definition: A **social choice function** is a function $C : L^n \rightarrow O$, where

- $N=\{1,2,..,n\}$ is a set of **agents**
- O is a finite set of **outcomes**
- L is the set of non-strict total orderings over O.

Definition: A **social welfare function** is a function $C : L^n \rightarrow L$, where N, O, and L are as above.

Notation:
We will denote ***i*'s preference order** as $\succeq_i \in L$, and a **profile** of preference orders as $[\succeq] \in L^n$.

Recap: Voting Scheme Properties

Definition:
W is **Pareto efficient** if for any $o_1, o_2 \in O$,

$$(\forall i \in N : o_1 \succ o_2) \implies (o_1 \succ_W o_2)$$

Definition:
W is **independent of irrelevant alternatives** if, for any $o_1, o_2 \in O$ and any two preference profiles $[\succ'], [\succ''] \in L$,

$$(\forall i \in N : o_1 \succ_i' o_2 \iff o_1 \succ_i'' o_2) \implies (o_1 \succ_{W[\succ']} o_2 \iff o_1 \succ_{W[\succ'']} o_2)$$

Definition:
W does not have a **dictator** if

$$\neg i \in N : \forall [\succ] \in L^n : \forall o_1, o_2 \in O : (o_1 \succ_i o_2) \implies (o_1 \succ_W o_2)$$

Recap: Arrow's Theorem

Theorem: (Arrow, 1951)
If $|O| > 2$, any social welfare function that is Pareto efficient and independent of irrelevant alternatives is dictatorial.

- Unfortunately, restricting to social choice functions instead of full social welfare functions doesn't help.

Theorem: (Muller-Satterthwaite, 1977)
If $|O| > 2$, any social choice function that is weakly Pareto efficient and monotonic is dictatorial.

Mechanism Design

- In the social choice lecture, we assumed that agents report their preferences **truthfully**
- We now allow agents to report their preferences **strategically**
- Which social choice functions are implementable in this new setting?

Bayesian Game Setting

Definition:
A **Bayesian game setting** is a tuple (N, O, Θ, p, u) where

- N is a finite set of n **agents**,
- O is a set of **outcomes**,
- $\Theta = \Theta_1 \times \ldots \times \Theta_n$ is a set of possible **joint type vectors**,
- p is a **common prior** distribution over Θ, and
- $u = (u_1, \ldots, u_n)$, where $u_i : O \rightarrow \mathbb{R}$ is the **utility function** for player i.

This differs from a Bayesian game only in that utilities are defined on **outcomes** rather than **actions**, and agents are not (yet) endowed with an action set.

Mechanism

Definition:
A **mechanism** for a Bayesian game setting (N,O,Θ,p,u) is a pair (A,M), where

- $A = A_1 \times \ldots \times A_n$, where A_i is the set of **actions** available to agent i, and
- $M : A \rightarrow \Delta(O)$ maps each **action profile** to a distribution over **outcomes**

Intuitively, a **mechanism designer** (sometimes called the **center**) needs to decide among outcomes in some Bayesian game setting, and so they design a mechanism that **implements** some social choice function.

Dominant Strategy Implementation

Definition:
Given a Bayesian game setting (N,O,Θ,p,u), a mechanism (A,M) is an **implementation in dominant strategies** of a social choice function C (over N and O) if for any vector u of utility functions,

1. The Bayesian game $(N, A, \Theta, p, u \circ M)$ induced by (A,M) has an equilibrium in dominant strategies, and

2. In any such equilibrium a^*, we have $M(a^*) = C(u(\cdot,\theta))$.

Bayes-Nash Implementation

Definition:
Given a Bayesian game setting $(N, O, \boldsymbol{\Theta}, p, u)$, a mechanism (A, M) is an **implementation in Bayes-Nash equilibrium** of a social choice function C (over N and O) if

1. There exists a Bayes-Nash equilibrium of the Bayesian game $(N, A, \boldsymbol{\Theta}, p, u \circ M)$ induced by (A, M) such that

2. for every type profile $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ and action profile $a \in A$ that can arise in equilibrium, $M(a) = C(u(\cdot, \boldsymbol{\theta}))$.

Direct Mechanisms

- The space of all functions that map actions to outcomes is impossibly large to reason about
- Fortunately, we can restrict ourselves without loss of generality to the class of **truthful, direct** mechanisms

Definition: A **direct** mechanism is one in which $A_i = \Theta_i$ for all agents i.

Definition:
A direct mechanism is **truthful** (or **incentive compatible**) if, for all type profiles $\theta \in \Theta$, it is a dominant strategy in the game induced by the mechanism for each agent to report their true type.

Definition:
A direct mechanism is **Bayes-Nash incentive compatible** if there exists a Bayes-Nash equilibrium of the induced game in which every agent always truthfully reports their type.

Revelation Principle

Theorem: (Revelation Principle)
If there exists any mechanism that implements a social choice function C in dominant strategies, then there exists a direct mechanism that implements C in dominant strategies and is truthful.

Revelation Principle Proof

1. Let (A,M) be a mechanism that implements C in Bayesian game setting (N,O,Θ,p,u).

2. Construct the **revelation mechanism** (Θ,M') as follows:

   - For each type profile θ, let $a^*(\theta)$ be the strategy profile in which every agent plays their dominant strategy in the game induced by (A,M).

   - Let $M'(\theta) = M(a^*(\theta))$.

3. Each agent reporting type $\hat{\theta}_i$ will yield the same outcome as every agent of type $\hat{\theta}_i$ playing their dominant strategy in M

4. So it is a dominant strategy for each agent to report their true type $\hat{\theta}_i = \theta_i$.

Exact same argument can be followed for Bayes-Nash incentive compatible direct implementation.

General
Dominant-Strategy Implementation

Theorem: (Gibbard-Satterthwaite)
Consider any social choice function C over N and O. If

1. $|O| > 2$ (there are at least **three** outcomes),

2. C is **onto**; that is, for every outcome $o \in O$ there is a preference profile $[\succ]$ such that $C([\succ]) = o$
(this is sometimes called **citizen sovereignty**), and

3. C is dominant-strategy **truthful**,

then C is **dictatorial**.

Hold On A Second

- Haven't we already seen an example of a dominant-strategy truthful direct mechanism?
- **Second Price Auction**
  - **Outcomes** are { (i gets object, pays \$$x$) | $i \in N$, $x \in \mathbb{R}$ }
  - **Types** are $\theta_i = \mathbb{R}$, where an agent i with type x prefers
    all outcomes where i gets object for ≤ \$$x$
    to all outcomes where i does not get x
    to all outcomes where i gets x for > \$$x$.
  - **Social choice function**: Assign the item to the agent with the highest type
  - **Actions**: Agents directly announce their type via sealed bid
- **Question:** Why is this not ruled out by Gibbard-Satterthwaite?

Restricted Preferences

- Gibbard-Satterthwaite only applies to social choice functions that operate on every possible preference ordering over the outcomes
- By restricting the set of preferences that we operate over, we can circumvent Gibbard-Satterthwaite

Quasilinear Preferences

Definition:
Agents have **quasilinear utility functions** (or **quasilinear preferences**) in an n-player Bayesian game setting when

1. the set of outcomes is $O = X \times \mathbb{R}^n$ for a finite set X,

2. the utility of agent i given joint type $\boldsymbol{\theta}$ for an element $(x,p) \in O$ is $u_i(o,\boldsymbol{\theta}) = v_i(x,\boldsymbol{\theta}) - f_i(p_i)$, where

3. $v_i : X \times \boldsymbol{\Theta} \rightarrow \mathbb{R}$ is an arbitrary function, and

4. $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$ is a monotonically increasing function.

Quasilinear Preferences, informally

- Intuitively: Agents preferences are split into
  1. finite set of **nonmonetary** outcomes (e.g., allocation of an object)
  2. monetary **payment** made to the center (possibly negative)
- These two preferences are **linearly** related
- Agents are permitted **arbitrary preferences** over nonmonetary outcomes, but **not over payments**
- Agents care only about the **outcome selected** and their **own payment**
- If every agent has linear utility for money with the same slope, then we are in the **transferrable utility** setting

Direct Quasilinear Mechanism

Definition:
A **direct quasilinear mechanism** is a pair (χ, p), where

- $\chi : \Theta \to \Delta(X)$ is the **choice rule**, which maps from a profile of reported types to a distribution over nonmonetary outcomes, and
- $p : \Theta \to \mathbb{R}^n$ is the **payment rule**, which maps from a profile of reported types to a payment for each agent.

Groves Mechanisms

Definition:
Groves mechanisms are direct quasilinear mechanisms (χ,p) for which

$$\chi(\hat{v}) = \arg\max_x \sum_i \hat{v}_i(x)$$

$$p_i(\hat{v}) = h_i(\hat{v}_{-i}) - \sum_{j \neq i} \hat{v}_j(\chi(\hat{v}))$$

- Groves mechanisms **implement** any social welfare maximizing choice function in **dominant strategies** (**why**?)

Vickrey-Clarke-Groves Mechanism

Definition:
The Vickery-Clarke-Groves mechanism is a direct quasilinear mechanism (χ,p), where

$$\chi(\hat{v}) = \arg\max_x \sum_i \hat{v}_i(x)$$

$$p_i(\hat{v}) = \sum_{j \neq i} \hat{v}_j(\chi(\hat{v}_{-i})) - \sum_{j \neq i} \hat{v}_j(\chi(\hat{v}))$$

- Each agent *i* pays the difference between the other agents' utility if *i* weren't there and the agents' utility now that *i* is there
    - Each agent pays their **externality**
- **Question:** Why don't we use this for **everything**?

Second Price Auctions *Are* VCG

- The **second price auction** is VCG in the single-item auction setting
- Object is awarded to agent with **highest valuation**; this maximizes the sum of agent valuations for the outcome
- Externality of **winning agent** is the value that **next-highest-valuation agent** could have gotten by winning the auction
- Externality of **losing agent** is nothing; if they weren't there, the outcome would be no different

Summary

- **Mechanism design**: Setting up a system for **strategic agents** to provide input to a **social choice function**
- **Revelation Principle** means we can restrict ourselves to **truthful direct** mechanisms without loss of generality
- Non-dictatorial dominant-strategy mechanism design is **impossible in general** (Gibbard-Satterthwaite)
- **Groves mechanisms** (especially **VCG**) implement any welfare-maximizing social choice function in dominant strategies for special case of **quasi-linear preferences**